Increasing the Understanding between a Dining Table Robot Assistant and the User*

Samuel Olatunji, Tal Oron-Gilad, Yael Edan

Abstract— **This study is a preparatory stage of a larger study intended to increase the understanding between a dining table robot assistant and the user. The users are expected to be older adults who need assistance in their daily lives but the study begins with investigating the level of understanding with younger adults with the intention of comparing the interaction with older adults in further studies. The aim of the experiment is to identify the most appropriate mode of communication from the robot which will convey the state of the interaction between the user and the non-humanoid robot. The results of the present study reveal that voice feedback from the robot aids better understanding of the state of interaction compared to visual feedback in the absence of background noise while the visual feedback aids better understanding in the presence of noise. Even though most of the users had an opaque understanding of the interaction with the robot while using the voice feedback mode, the results point to the possibility of obtaining better understanding if both feedback modes are combined, to highlight the advantage of each modality, and the content of the information provided is improved. The study is the initial step towards a design framework for improving the understanding between a socially assistive robot (such as a table setting robot) and the user.**

I. Introduction

Socially assistive robots (SARs) are a possible solution to bridge the elder care gap [1], which is defined as the dearth of caregivers and healthcare professionals available to cater for older adults [2]. SARs can assist older adults in some activities of daily living such as meal setting [3]–[5]. This constitutes a form of human-robot interaction (HRI) where older adults are expected to interact with a robot serving as a dining table robot assistant. One of the challenges involved in this interaction which this study intends to address, is the mismatch commonly observed in the user's understanding of the state of the robot relative to the robot's actual state. This mismatch could lead to misuse – if the user over-relies on the robot, or disuse – if the user under-utilizes the robot [6]. In the sensitive setting of elder care, such consequences can significantly degrade the quality of user-robot interaction. The research addresses the following question: which information presentation mode from a non-humanoid table setting robot effectively communicates the state of the interaction to the user?

II. Literature Review

Optimal robot performance and user experience during human robot interaction (HRI) are important aspects that define quality of interaction [7]. Understanding the robot's state is a crucial link in the metrics of assessments which needs to be taken into consideration [8]. Understanding in the context of HRI can be described as the extent to which a human and a robot have adequate knowledge about each other's state to be able to successfully interact with each other [9]. Communicative actions could be sent from the user to the robot or vice versa in form of instructions or feedback [10]. These communicative actions when presented in the most comprehensible form promotes understanding which leads to a successful interaction of the user with the robot [8], [11]. It is a form of bidirectional presentation of information where the instructions could originate from the user or robot, encoded in a specific mode or multimode (such as visual, audial or gestural) and decoded through various mode recognition or perception techniques (such as GUI, speech or gesture recognition mechanisms) [12]. This bidirectional communication keeps both parties aware of the factors underlying each other's actions and allows them to correct erroneous factors that each may have [13]. Successful bidirectional communication between the robot and the human supports transparency of the interaction, team performance and trust in the automation [13]. The extent to which the human understands the robot's communicative action can be referred to as states of understanding as used by Clark and Schaefer [14] and further elaborated by Doran et al. [11] as presented in Table I.

TABLE I. States of Understanding

States of Understanding	Description
Opaque	Recipients perceive the inputs and outputs of a system without knowledge of how the input is mapped to the output.
Interpretable	Recipients perceive not just the inputs and outputs of a system but can also observe all the details that produced the output from the input. Understanding the details that map the input to the output usually requires the user to have background knowledge of the data and domain.
Comprehensible	Recipients perceive the inputs and outputs of a system and can also comprehend the meaning and relationship between the input and output. Symbols and words are often encoded in the system with a knowledge base that can help the user relate the input to the output.

Several studies have explored different modalities through which a robot may express its state to the user. These modalities include buzzers, light projections, motion [15], gestures, facial expressions, body language [16], speech [17] and augmented reality [18]. The choice of modality to use is strongly predicated on the several factors which particularly includes the type and capability of the robot [15], context of use and noise conditions [7], [11]. Noise has been observed in

* This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 721619 for the SOCRATES project.

All authors are from the Department of Industrial Engineering and Management, Ben-Gurion University of the Negev Beersheba 84105, Israel (phone: +972-86461434; email: olatunji@post.bgu.ac.il, orontal@bgu.ac.il, yael@bgu.ac.il).

existing studies to corrupt the accurate interpretation and comprehension of information communicated to recipients [9]. This study hypothesizes that there will be an interaction between the mode of feedback and background noise. A second hypothesis is that visual feedback will influence a higher level of understanding in the presence of background noise, while voice feedback will do so in quiet environments. Extensive user studies are required to explicitly identify the most appropriate mode of communication that will promote understanding in the case of socially assistive robots that have no semblance of human morphological features such as the meal setting robotic arm used in this study.

III. METHODS

A. Overview

There are four groups in the study. The groups consisted of different combinations of feedback modes and noise. The feedback was provided by the robot to give the user information on the status of the interaction while the noise was simulated to depict typical noisy settings. Participants were asked to give voice commands to the robot to perform a pick and place task similar to what would be required in setting utensils and food items on a dining table. Objective and subjective measures were taken to assess the understanding the users had regarding the state of the interaction based on the feedback given by the robot. The overall experience with the robot was also assessed. The study took place at the intelligent robotics laboratory, Ben-Gurion University of the Negev, Israel.

B. Apparatus

The dining table robot assistant used was a robotic arm – KUKA iiwa (Intelligent Industrial Work Assistant) LBR (Lightweight Robot) with seven DOF (Degrees of Freedom). The KUKA enables fast development and integration of devices, using Robot Operating System (ROS) [19]. It is a lightweight robot for industrial applications that is designed for safe close cooperation between human and robot on highly sensitive tasks [20].

C. Participants

A convenience sample of sixteen people participated in the experiment (6 Females, 10 Males) aged 21-57 (mean 29.2 years). The intention is to experiment first with younger people who are more readily accessible and then proceed to use the lessons learned for the experiment involving older adults. There were 8 participants with Engineering background while the other 8 were from other disciplines. Each participant completed the study separately at different timeslots, so there was no contact between participants.

D. Experimental Design

The experiment was set as a between-participant factorial design with manipulations of feedback and noise conditions as independent variables. The feedback modes used were voice and visual feedback modes while the noise manipulation was a condition with the presence of an alarm noise in the background and without it, as illustrated in Table II. Participants were assigned randomly to one of the four groups. Each participant had either voice or visual feedback in the presence or absence of noise based on the group assigned. The visual feedback was in the form of a display on a screen situated near the robot displaying 'Good Work' on a green background when the robot sensed the voice command given by the participant and was moving as commanded. The display showed 'Not Done' on a red background when the robot was yet to carry out the commanded task or could not carry out the commanded task. The feedback information stayed on the screen till the next command was issued and the next feedback information related to the new command was displayed. The voice feedback gave the same information but in the form of a simulated human voice which was given repeatedly at specific intervals till the next command was issued. The noise effect was implemented in the form a repetitive rhythmic alarm sound in the background at approximately 55dB. The alarm was switched off in the groups without noise, and the sound level in the lab was maintained at approximately 35dB. The sound level of the voice feedback was at approximately 60dB such that the participants could hear the voice feedback well above the alarm noise.

TABLE II. EXPERIMENTAL GROUPS

		Alarm Noise	
		Present	***Absent***
Feedback	**Voice Feedback**	Group A	Group B
	Visual Feedback	Group C	Group D

E. Experimental task

Participants were assigned a task which consisted of two trials: The first trial was to give voice commands to lead the robot to pick and place pre-arranged fruits into a bowl while the second trial was to give voice commands to the robot to pick cups and arrange them in a predefined configuration (*Fig. 1*). The trials were counterbalanced between participants. It was designed using a Wizard-of-Oz technique where the users' commands were translated to the robot's motion in real time via the keypad of the robot by an experimenter.

F. Procedure

At the start of the experiment, the participants were asked to fill a consent form which described the experiment and what the participant was required to do. The participants were then asked to complete a pre-test questionnaire which included some demographic information, a Technology Adoption Propensity (TAP) index [21] and a Negative Attitude toward Robots Scale (NARS) [22]. The robot was then introduced to the participants as their table setting robot assistant who could carry out their commands to set items on the dining table. An instruction set of 8 commands was given to the participants to control the robot as described in Table III. Participants were asked to command the robot to accomplish the two trials described in the experimental design. Post-trial questionnaires were administered after each trial and a final questionnaire at the end of the experiment to assess the subjective experience with the robot assistant.

Fig. 1: Experimental setup using the KUKA robot

TABLE III. SET OF COMMANDS TO CONTROL THE ROBOT

Command	***Action of the robot***
Left	Moves towards the negative x axis
Right	Moves towards the positive x axis
Forward	Moves towards the positive y axis
Backward	Moves towards the negative y axis
Up	Moves towards the positive z axis
Down	Moves towards the negative z axis
Open	Opens the gripper
Close	Closes the gripper

IV. RESULTS AND DISCUSSION

The results obtained from the objective and subjective measures are presented in the following subsections.

A. Demographics

There was an equal distribution of participants within the 4 groups. The participants were mostly acquainted with the use of innovative technologies. On a scale of 1 (strongly disagree) to 5 (strongly agree), the TAP index reveals that most of the participants are optimistic about technology providing more control and flexibility in life (mean = 4.09, SD = 0.86). The NARS reveals that the participants do not have negative feelings about situations in which they interacted with a robot (mean = 2.14, SD = 1.2).

B. Objective Measures

The objective measures were the average time it took participants to complete the task (in seconds) and the average number of errors made by the robot while being commanded to pick and place the items. The independent variables for the experiment were the manipulations of the feedback mode and presence of noise. The dependent variable is the level of understanding the user has regarding the state of the interaction.

The average time it took participants to complete the task (consisting of both trials) in the experiment was 369 seconds (SD = 82 seconds). In the presence of the background noise, participants with visual feedback (group C) spent the shortest time on the tasks (mean = 327 seconds, SD = 16.84 seconds). In the absence of the background noise, participants with voice feedback (group A) spent shorter time on the tasks than participants with visual feedback (mean= 364 seconds, SD = 30.01 seconds). This is presented in Fig. 2. It is assumed that the longer it took the participants to complete the tasks, the less understanding they had regarding the interaction based on the feedback given by the robot.

Participants with the voice feedback in the absence of noise (Group B) experienced the least number of errors (mean =1, SD = 0.82) while participants with visual feedback in the presence of noise encountered the highest number of errors (mean = 2, SD = 1.63). The error values are indicated in Fig. 2 (in purple). It is assumed that less errors indicated to some extent that the participants had a good understanding of the interaction based on the feedback provided by the robot.

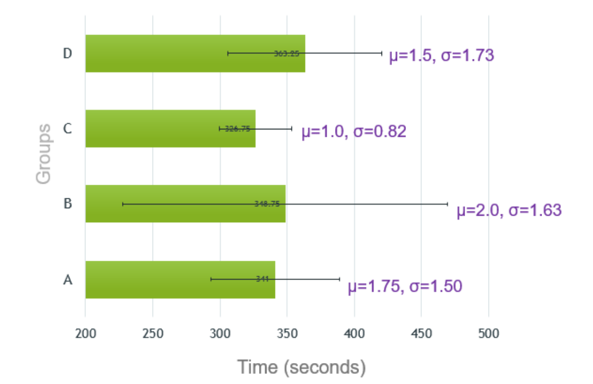

Fig. 2: Average time it took participants in each group to complete a task (bars represent SE), purple number represents average number of errors per task and SD.

B. Subjective Measures

The experience of interacting with the table setting robot is presented in *Fig. 3*. Only 2 (13%) of the participants considered the robot as understandable. These were in the groups with voice feedback. The subjective rating of the level of understanding the users in each of the groups have regarding the state of the interaction is presented in Fig. 4. The groups with voice feedback had more participants who understood the robot's feedback at an opaque level. There is a high possibility that their level of understanding was affected by the presence of noise since there were some participants with voice feedback in the absence of noise whose subjective ratings indicate a comprehensive understanding of the information the robot was communicating.

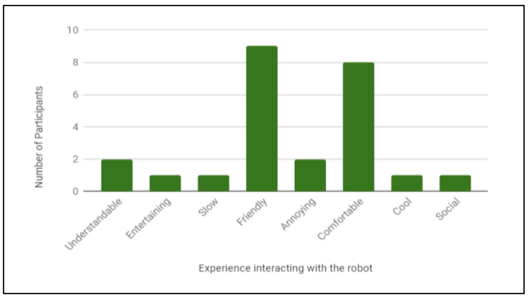

Fig. 3: Users' perception of the robot

Both the objective and subjective results reveal that visual feedback from the robot aided a better understanding of the state of interaction compared to voice feedback in the presence of background noise whereas participants experienced better understanding with the voice feedback when the noise was absent. Both feedback modes can therefore be combined to create an improved communication mode rather than utilizing voice feedback as the only communication mode.

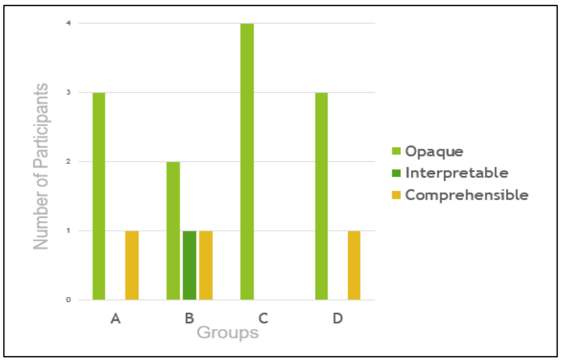

Fig. 4: Number of ratings of level of understanding for each group

The subjective experience of the participants revealed that 75% of the participants have just an opaque understanding of the interaction, despite their positive TAP scores and irrespective of the feedback mode being used. This therefore brings to the fore, the possibility that the content of information being displayed or spoken in words may have been insufficient to convey a comprehensive level of understanding of the information being presented by the robot. Three levels of information content could be displayed or voiced out which are connected with presenting what the robot is doing, the reason for the action(s) and consequence(s) of the action(s) [23]. In this study, only the state of the interaction (level 1) was displayed. Future work to improve the understanding will entail varying the content of the feedback to include the reason for the robot's actions (level 2) and the consequences of such actions (level 3). These studies will also be conducted with more participants to provide sufficient data for standard statistical significance tests.

V. CONCLUSION

The study revealed that the voice feedback mode used in the interaction between a table setting robot assistant and the user aided better understanding of the interaction state compared to the visual feedback in the absence of background noise. Visual feedback provided better understanding than voice feedback when noise is present. This gives insight for the next stage of the research which would include testing the combination of both feedback modality modes and varying the content of the information being provided to further improve the user's level of understanding.

REFERENCES

[1] D. Tang, B. Yusuf, J. Botzheim, N. Kubota, and C. S. Chan, "A novel multimodal communication framework using robot partner for aging population," *Expert Syst. Appl.*, vol. 42, no. 9, pp. 4540–4555, 2015.

[2] N. Super, "Who Will Be There to Care? The Growing Gap between Caregiver Supply and Demand," *Natl. Heal. Policy Forum*, 2002.

[3] D. McColl, W. Y. G. Louie, and G. Nejat, "Brian 2.1: A Socially assistive robot for the elderly and cognitively impaired," *IEEE Robot. Autom. Mag.*, vol. 20, no. 1, pp. 74–83, 2013.

[4] C. A. Smarr *et al.*, "Domestic Robots for Older Adults: Attitudes, Preferences, and Potential," *Int. J. Soc. Robot.*, vol. 6, no. 2, 2014.

[5] C. A. Smarr, A. Prakash, J. M. Beer, T. L. Mitzner, C. C. Kemp, and W. A. Rogers, "Older adults' preferences for and acceptance of robot assistance for everyday living tasks," *Proceedings of the Human Factors and Ergonomics Society*. pp. 153–157, 2012.

[6] R. Parasuraman and V. Riley, "Humans and Automation: Use, Misuse, Disuse, Abuse," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 39, no. 2, pp. 230–253, 1997.

[7] S. Bensch, A. Jevtić, and T. Hellström, "On Interaction Quality in Human-Robot Interaction," *Proc. 9th Int. Conf. Agents Artif. Intell. - Vol. 1 ICAART*, vol. 1, pp. 182–189, 2017.

[8] N. Balfe, S. Sharples, and J. R. Wilson, "Understanding Is Key: An Analysis of Factors Pertaining to Trust in a Real-World Automation System," *Hum. Factors*, no. 1983, 2018.

[9] T. Hellström and S. Bensch, "Understandable Robots -," *Paladyn, J. Behav. Robot*, 2018.

[10] N. Mirnig, S. Riegler, A. Weiss, and M. Tscheligi, "A case study on the effect of feedback on itinerary requests in human-robot interaction," *Ro-Man, 2011 Ieee*, pp. 343–349, 2011.

[11] D. Doran, S. Schulz, and T. R. Besold, "What Does Explainable AI Really Mean? A New Conceptualization of Perspectives," 2017.

[12] N. Mirnig, A. Weiss, and M. Tscheligi, "A communication structure for human-robot itinerary requests," *Human-Robot Interact. (HRI), 2011 6th ACM/IEEE Int. Conf.*, pp. 205–206, 2011.

[13] J. Y. C. Chen, S. G. Lakhmani, K. Stowers, A. R. Selkowitz, J. L. Wright, and M. Barnes, "Situation awareness-based agent transparency and human-autonomy teaming effectiveness," *Theor. Issues Ergon. Sci.*, vol. 19, no. 3, pp. 259–282, 2018.

[14] H. H. Clark and E. F. Schaefer, "Collaborating on contributions to conversations," *Lang. Cogn. Process.*, vol. 2, no. 1, pp. 19–41, Jan. 1987.

[15] K. Kobayashi and S. Yamada, "Making a Mobile Robot to Express its Mind by Motion Overlap," *Adv. Human-Robot Interact.*, no. December 2009, pp. 342–356, 2009.

[16] H. Knight and R. Simmons, "Layering Laban Effort Features on Robot Task Motions," *Proc. Tenth Annu. ACM/IEEE Int. Conf. Human-Robot Interact. Ext. Abstr. - HRI'15 Ext. Abstr.*, pp. 135–136, 2015.

[17] N. Mavridis, "A review of verbal and non-verbal human-robot interactive communication," *Rob. Auton. Syst.*, vol. 63, no. P1, pp. 22–35, 2015.

[18] J. Carff, M. Johnson, E. M. El-Sheikh, and J. E. Pratt, "Human-robot team navigation in visually complex environments," *2009 IEEE/RSJ Int. Conf. Intell. Robot. Syst. IROS 2009*, pp. 3043–3050, 2009.

[19] S. Mokaram *et al.*, "A ROS-integrated API for the KUKA LBR iiwa collaborative robot," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 15859–15864, Jul. 2017.

[20] M. A. K. Bahrin, M. F. Othman, N. H. N. Azli, and M. F. Talib, "Industry 4.0: A review on industrial automation and robotic," *J. Teknol.*, vol. 78, no. 6–13, pp. 137–143, 2016.

[21] M. Ratchford and M. Barnhart, "Development and validation of the technology adoption propensity (TAP) index," *J. Bus. Res.*, vol. 65, no. 8, pp. 1209–1215, 2012.

[22] D. S. Syrdal, K. Dautenhahn, K. Koay, and M. L. Walters, "The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study," *23rd Conv. Soc. Study Artif. Intell. Simul. Behav. AISB*, pp. 109–115, 2009.

[23] J. Y. C. Chen, K. Procci, M. Boyce, J. Wright, A. Garcia, and M. J. Barnes, "Situation Awareness – Based Agent Transparency," no. April, pp. 1–29, 2014.